DISCRIMINATION IN AUDITORY AND VISUAL PATTERNS

By Jo Ann S. Kinney, U.S.N. Medical Research Laboratory

Within the realm of audition, there are relatively few data on the perception of complex sounds other than speech, and few laws on the perception of form and pattern with which to compare the laws of visual organization. Recent contributions to the field have been made by Miller and Heise in two experiments on auditory patterns.[1] In the first of these, they found that two alternating tones of different pitch will be heard as one pattern or trill if the separation in frequency between the tones is not too great. Beyond a certain size of separation, however, the trill breaks, and two unrelated tones are heard. Heise and Miller later experimented with other types of auditory pattern and found a similar phenomenon. A single variable tone, presented with others in a pattern, becomes isolated from the rest of the pattern when the frequency-separation is increased beyond a certain point. The "single-tone 'figure' is heard as an isolated 'pop' . . . with the onrushing steam of melodic pattern in the background."[2]

Temporal variation was not studied by Miller and Heise, but it is of obvious importance in auditory perception, since auditory stimuli are patterned in time. It does follow from their results, however, that if a tone is 'isolated' from the rest of an auditory pattern by a large frequency-separation, its temporal position with respect to the rest of the pattern should be more difficult to judge. Temporal discrimination therefore could be used to measure which tones are included in a perceptual group.

A further advantage resulting from the use of temporal discrimination as a measure of auditory patterning is that the results may be compared with data of visual patterning by means of the spectrographic transformation between vision and audition. Time is one of the major axes in this transformation, being comparable to the horizontal dimension of visual space, while pitch becomes the vertical dimension of space. There are a number of instances in the literature indicating that the spectrographic transformation preserves the patterning or organization of stimuli be-

* Received for publication June 15, 1960. The auditory portion of this paper is adapted from a Doctoral dissertation submitted to the University of Connecticut. The work was supported, in part, by the Haskins Laboratories in connection with Contract AF 19(604)-2285, which is monitored by Operational Applications Laboratory, AFCRC, Bedford, Mass. The author is indebted to Dr. Alvin M. Liberman who directed the research.

[1] G. A. Miller and G. A. Heise, Trill threshold, *J. acoust. Soc. Amer.*, 22, 1950, 637-638; An experimental study of auditory patterns, this Journal, 64, 1951, 68-77.

[2] *Op. cit.*, this Journal, 72.

tween the two senses. The first such indication came from the development of the Sound Spectrograph by Bell Telephone Laboratories and from the comparison of spectrographic and oscillographic displays of speech. Some individuals, after training, could read spectrograms of speech, a feat impossible with the oscillograph.[3]

The reverse of the Sound Spectrograph, the Pattern Playback developed by Haskins Laboratories, converts visual displays into sounds.[4] This instrument has been used successfully in determining for the first time the acoustic cues for speech, further attesting to the effectiveness of the transform in preserving the organization of the stimuli between the two senses.[5]

Studies of Licklider, Bindra, and Pollack on the intelligibility of speech,[6] and by Cooper, Liberman, and Borst on pattern perception,[7] also indicate the parallelism that exists between visual and auditory perception when the spectrographic transform is used. Heise and Miller conclude that their results with sound-patterning are in agreement with the laws of visual organization and that "the shape of the auditory pattern and the threshold for the integration of the variable tone into the pattern are approximately what one would expect from corresponding visual figures, if the frequency and time coordinates of the auditory figure are replaced by vertical and horizontal spatial coordinates, respectively."[8]

Other instances could be cited, the most obvious being that musical notation depends on this device and has been read successfully for centuries. It should be pointed out, however, that there is no indication that this transformation is the only one, or that it will work perfectly; it is simply, on the weight of the evidence, the most useful one we have at present.

This experiment then was designed, first, to provide quantitative data on the perception of temporal differences in auditory patterns and, secondly, to compare the auditory results with similar data from vision. Specifically, the effect of two major variables on auditory temporal discrimination was investigated. The frequency-separation of the elements or tones in the patterns was chosen as one variable, to see if the dependence of perceptual grouping upon pitch-relationships, found by Miller and Heise, could be repeated for temporal discrimination. The type of pattern, or the manner in which the tones are combined, was chosen as the second parameter, since it relates most obviously to form-perception in all modalities.

[3] R. K. Potter, Visible patterns of sound, *Science*, 102, 1945, 463-470.

[4] F. S. Cooper, Spectrum analysis, *J. acoust. Soc. Amer.*, 22, 1950, 761-762; A. M. Liberman, Pierre Delattre, and F. S. Cooper, The role of selected stimulus-variables in the perception of the unvoiced stop consonants, this JOURNAL, 65, 1952, 497-516.

[5] F. S. Cooper, P. C. Delatttre, A. M. Liberman, J. M. Borst, and L. J. Gerstman, Some experiments on the perception of synthetic speech sounds, *J. acoust. Soc. Amer.* 24, 1952, 597-606; P. C. Delattre, F. S. Cooper, A. M. Liberman, and L. J. Gerstman, Speech synthesis as a research technique, *Proc. VIIth International Congress of Linguists* (1952), London, 1956, 545-561; A. M. Liberman, Some results of research on speech perception, *J. acoust. Soc. Amer.*, 29, 1957, 117-123.

[6] J. C. R. Licklider, D. Bindra, and I. Pollack, The intelligibility of rectangular speech-waves, this JOURNAL, 61, 1948, 1-20.

[7] Cooper, Liberman, and Borst, The interconversion of audible and visible patterns as a basis for research in the perception of speech, *Proc. nat. Acad. Sci.*, 37,1951, 318-325.

[8] Heise and Miller, *op. cit.*, this JOURNAL, 76.

To obtain quantitative visual data for comparison, the spectrographic transform was used to convert the auditory patterns into visual patterns. Since there is as yet no evidence on the appropriate scale to use in such a conversion of axes, a linear transformation was arbitrarily chosen, with the result that the patterns were the same when the linear coördinates of frequency and time were replaced by linear dimensions of length. Thus, for visual experimentation, the elements of the patterns are rectangles, and the discriminative task is that of judging the spacing between rectangles in the horizontal dimension. The two experimental variables are the amount of separation between elements in the vertical dimension and the type of pattern.

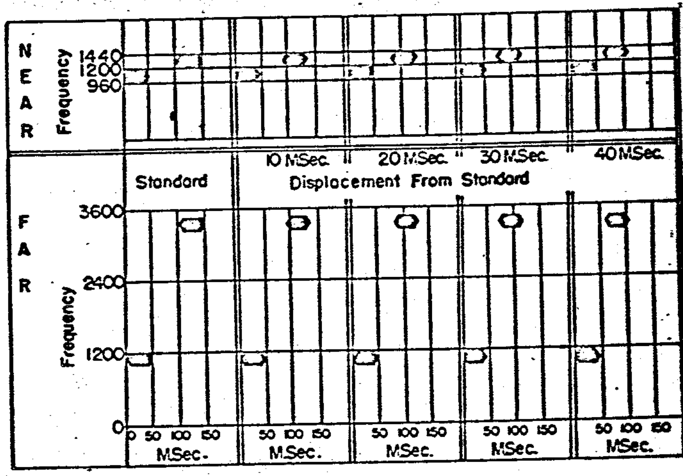

FIG. 1. SCALE-DRAWING OF THE TWO FREQUENCY-CONDITIONS OF THE BASIC ELEMENTS FOR THE STANDARD AND VARIABLES IN THE AUDITORY PATTERNS.

PART I: AUDITION

Apparatus. The stimuli were produced by the Pattern Playback, the instrument designed to convert spectrographic displays into sound. Spectrograms are painted in white on transparent material, placed on the moving apparatus, and passed through a modulated light beam which has a fundamental frequency of 120 ~ and 50 harmonics. The light reflected from the white paint is converted into the corresponding frequencies of sound by a phototube, amplifier, and speaker-system.

In this experiment, drawings were made of the desired patterns, sounds were generated from them by the Playback, and recordings of the sound were made on magnetic tape. The individual patterns were then placed in the desired orders for presentation to the *Os* by marking and splicing the magnetic tape. The reels of magnetic tape were played for the *Os* on a Magnecorder and loud-speaker system in a room specially treated to absorb sound reflections.

Stimuli. The basic elements in each pattern of sound consisted of two tones of different frequency. The tones were each of 50 m.sec. duration and were separated in the standard condition by 50 m.sec. of silence. In the variable conditions, the silence was of 10, 20, 30, or 40 m.sec. duration.

Two degrees of frequency-separation were used; one small and the other large. The former will be referred to as the 'near' condition and the latter as the 'far.' In the near-condition, the two tones were of 1200 and 1440 ~; in the far-condition, they were 1200 and 3480 ~.

The individual tones were not pure but consisted of two adjacent harmonics of

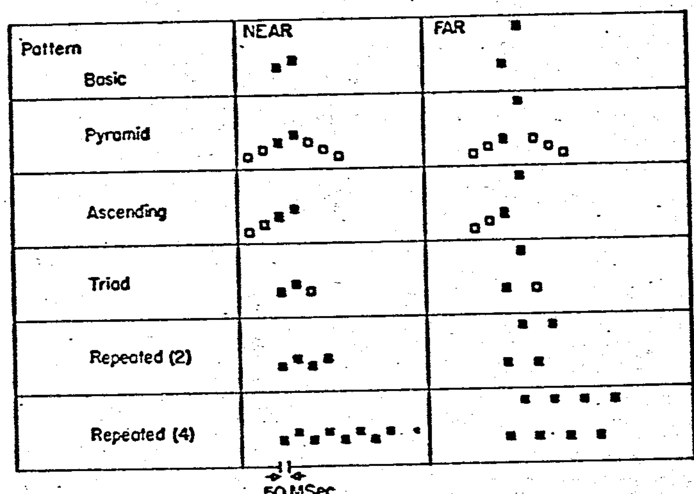

FIG. 2. SCHEMATIC DIAGRAM OF THE STIMULUS-PATTERNS

the 120 ~ fundamental; in specifying the frequency-level, the higher harmonic is cited. They started and ended gradually to avoid a burst of noise at onset and offset.

Fig. 1 presents the two frequency-separations of the basic elements for the standard and variable conditions. This is a scale-drawing of the patterns actually used and painted for conversion into sounds. Two such drawings were made of each pattern. One of the patterns used in the experiment consisted of these basic elements presented alone. For each of the other patterns, exactly the same basic elements, drawn in the same way, either were imbedded in a series of other tones or were repeated a number of times by themselves.

Six different patterns were used. Fig. 2 shows a schematic diagram of these patterns, with the basic elements represented by filled squares and the other tones by open squares. Whenever tones other than the basic elements were added, the tones were of the same length and shape as the basic elements, and 50 m.sec. of silence was interposed between them.

This diagram presents only the standard condition of 50 m.sec. of silence between the basic elements. For each pattern, the variable conditions, with 10-40 m.sec. of silence between the basic elements, were exactly the same as those presented in Fig. 1. Thus, it was always the highest tone in each pattern that was displaced temporally away from the standard condition in which the spacings between tones always were equal; all other tones remained in the same position on the time-scale. This means that, for those patterns which had a tone following the basic elements, a given displacement from the standard condition narrowed the silence between basic elements and widened it following the variable tone. For example, at 40 m.sec. displacement, the silent-time was 10 m.sec. between the basic elements and 90 m.sec. between the displaced tone and the one following it.

Presentation of stimuli. The *ABX* method was used to measure the *O*'s ability to discriminate the temporal separation between the elements. This method presents three patterns in a series. The first two are different, one being the standard and the

TABLE I

THE NUMBER OF *Os* TESTED ON EACH REEL

Os	No. sessions	Patterns presented: basic	pyramid	triad	ascending	repeated: two	repeated: four
Graduates	2	18	18	—	—	—	—
Undergrads	2	32	32	—	—	—	32
Undergrads	2	31	31	31	31	—	31
Undergrads	1	8	8	8	8	—	8
Undergrads	1	43	43	43	—	43	43
Columbia	1	6	6	6	6	—	6
Sailors	1	—	—	—	—	40	—
Total		138	138	88	45	83	120

other, a variable; the third pattern in the series is exactly the same as either the first or the second. The *O*'s task, therefore, is to decide whether the third pattern, *X*, is like, the first *A*, or the second, *B*.

The standard, with 50 m.sec. of silence between the basic elements, was compared with each of the variables that had 10-40 m.sec. of silence between elements. All possible orders were used for each comparison—*ABB*, *ABA*, *BAA*, and *BAB*. To control for the possibility that variations in the drawing or recording of the stimuli might serve as clues to the correct answer, the two identical temporal variations in each *ABX* series always were made from different drawings and recordings of the same pattern. A complete test for any one pattern consisted of all four orders for each of the four displacements, or 16 *ABX* series. These 16 series were placed on one reel of magnetic tape with the order of presentations randomized. Two frequency-conditions for each of the 6 patterns gave 12 reels.

The *O* was told that the difference between the patterns was of a temporal or rhythmical nature and to listen carefully for this aspect. He was further told that some of the differences might be difficult to detect and asked, where he was unsure, to guess. Each *O* was given data sheets on which to write down his answers.

These sheets were scored for the number of correct judgments at each of the temporal displacements.

The *Os* were, for the most part, undergraduates at the University of Connecticut, but they also included some graduate students from Columbia University, and enlisted men of the U. S. Navy. They were tested in small groups of less than 20 and heard as many reels as possible during the allotted testing time. All reels were not played to all *Os*, since an individual session, at most 45 min. long, was never long enough to encompass all of the variations; but the same *Os* always were tested with both the near and far frequency-conditions of a single pattern. The order of presentation of reels given within a single session was randomized, and each small

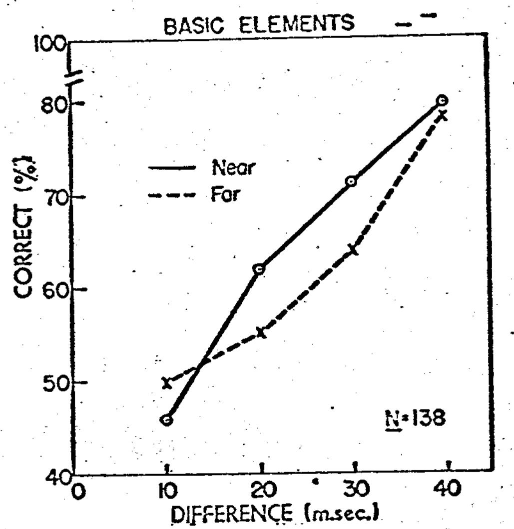

FIG. 3. RESULTS FOR THE TWO FREQUENCY-CONDITIONS OF THE BASIC ELEMENTS

group of *Os* was assigned a different order. The *Os* usually were asked to return for a second session a week later, and the same reels were repeated, in a different order, for those who did return. Table I gives all the information on the various patterns to which each group of *Os* was exposed and the total number of *Os* who heard each reel.

Results. The results obtained with the two frequency-separations of the basic elements are presented in Fig. 3, in terms of the average percentage correct for 138 *Os* at each temporal displacement. Normal psychophysical functions are found relating discrimination to the size of the difference to be discriminated. Both curves start with a chance-level of discrimination at 10 m.sec. and show increasingly larger percentages as the length of the temporal displacement is increased. The liminal values for both curves,

at the 75% point, are 30 -40 m.sec. The near-separation yields better discrimination throughout the curve except at the 10-m.sec. interval. This same increase in discrimination in the near-condition is found in the results for all of the other patterns. These data are given in Figs. 4 and 5;

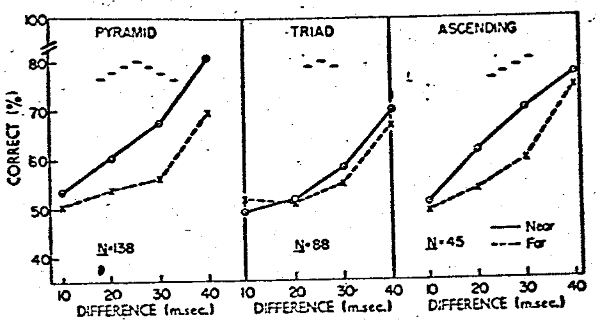

FIG. 4. RESULTS FOR THE TWO FREQUENCY-CONDITIONS OF THE PYRAMID, TRIAD, AND ASCENDING PATTERNS

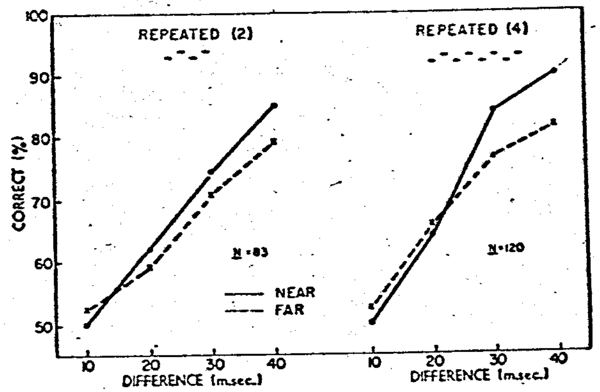

FIG. 5. RESULTS FOR THE TWO FREQUENCY-CONDITIONS OF THE REPEATED PATTERNS

wherever the results are above the chance-level, the near-condition yields better discrimination than the far.

It is obvious also from these data that discrimination varies with the type of pattern presented even where the actual temporal differences to be discriminated are identical. As a further check on this finding, comparisons were made only for *Os* who responded to all the reels in question; the purpose, of course, was to eliminate the effects of individual differences in sensitivity. In the previous comparison between frequency-separations, each *O* always was given both frequencies of a single pattern, which meant that individual differences already were eliminated. The results of these comparisons between patterns revealed consistent differences for all groups. The repeated patterns always showed discriminative functions superior to those of the other patterns, while the triad gave the poorest

TABLE II

THE DIFFERENCES BETWEEN MEAN RESULTS FOR THE BASIC ELEMENTS AND THE OTHER PATTERNS

Pattern	Mean	Mean diff. (Basic—Other)	*t*	P
Basic	10.39			
Pyramid	10.43	−0.04	−0.16	—
Ascending	10.18	0.21	0.74	—
Triad	9.10	1.29	4.80	1%
Repeated	11.91	−1.52	−4.95	1%

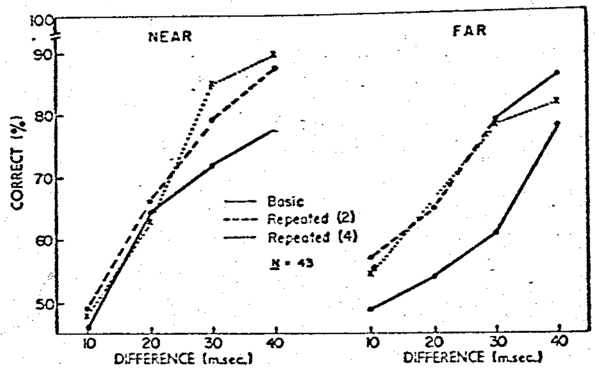

FIG. 6. EFFECT OF REPEATING THE BASIC ELEMENTS ON DISCRIMINATION

functions. The *t*-tests in Table II are given as an example of these differences; comparisons are made between the basic elements and four other patterns for a group of 31 *O*s, all of whom responded to all five patterns.

Data on the repeated patterns for a group of 43 *O*s, all of whom had the basic elements and two and four repetitions of the basic elements are plotted in Fig. 6. In both the near- and the far-conditions, there is better discrimination for the repeated patterns than for the basic elements alone. It should be noted also that the number of repetitions makes little difference in the discriminative functions, the curves for two repetitions and

for four repetitions intertwining at both near- and far-frequencies. Table III gives the results of *t*-tests performed on these data. No significant differences exist between two and four repetitions of the basic elements, while the differences between the basic elements and its repetitions all are significant.

PART II: VISION

The visual patterns used in this experiment were drawn in black India ink on 3 × 5-in. white cards. For presentation to the *O*s, they were placed in a holder against a large, white, Bainbridge-board screen. The screen

TABLE III

DIFFERENCES BETWEEN THE MEAN RESULTS FOR THE BASIC ELEMENTS, TWO, AND FOUR REPETITIONS

Frequency	* Patterns	Mean diff.	*t*	P
Near	Basic *vs.* two	0.86	1.82	5%
	Basic *vs.* four	1.00	2.23	1%
	Two *vs.* four	−0.12	−0.34	—
Far	Basic *vs.* two	1.84	3.60	1%
	Basic *vs.* four	1.56	3.45	1%
	Two *vs.* four	0.28	0.92	—

* The value given is for a one-tailed test, on the assumption that the repetitions always should be better than the elements alone.

was placed at the end of a 30-ft. alley which was painted a neutral grey and evenly illuminated to a level of 14 ft.-c.

Stimuli. The basic elements in each visual pattern were two solid rectangles, $\frac{1}{4} \times \frac{1}{8}$ in. The horizontal space between the two rectangles varied from $\frac{1}{4} - \frac{1}{8}$ in. in $\frac{1}{32}$-in. steps. At the viewing distance of 26 ft. 10 in., these steps gave separations of $2\frac{2}{3}$, $2\frac{1}{3}$, 2, $1\frac{2}{3}$, and $1\frac{1}{3}$ min. of visual angle. This variation in the horizontal dimension was used to measure discrimination.

Two degrees of spatial separation in the vertical dimension were used. In one, the 'near-condition,' the top of the first rectangle was in line with the bottom of the second. In the other, the 'far-condition,' $1\frac{1}{16}$ in. or 11.3 min. separated the two rectangles. A scale-drawing of the variations in the spacing of the basic elements is given in Fig. 7.

Five patterns were used: the basic elements, two and four repetitions of the basic elements, the triad, and the pyramid. The schematic drawing of the auditory patterns in Fig. 2 is equally applicable to the visual patterns, except that the 50-m.sec. interval is $\frac{1}{4}$ in. in the visual patterns. When other rectangles were added to the basic elements, a constant, horizontal spacing of $\frac{1}{4}$ in. was used, with only the top rectangle changed in spacing in relation to the others. Thus, as the top rectangle was moved to the left, the distance between it and the rectangle on its right was increased proportionately. Each of the variations of the patterns was drawn on its own 3 × 5-in. card. Two drawings were made of each variation.

Presentation of stimuli. The *ABX* method was used as in the auditory part. *O* was presented with three cards in succession and asked whether the third pattern was like the first or the second. The three cards varied in the amount of horizontal displacement between the basic elements.

Preliminary observations showed that *O*'s discrimination of the difference in spacing was extremely acute. No standard, therefore, was used, as was the case in the auditory problem, but rather each variation in spacing was combined with every other variation to obtain as fine a measure as possible. Thus, the 8/32-in. spacing between elements was combined in *ABX* series with 7/32 in.; 7/32 in. was combined with the 6/32 in.; 6/32 in. with 5/32 in.; and so forth. All such combinations, in which the difference to be discriminated was 1/32 in. (1/3 min.), are called the one-step differences. All combinations of 2/32-in. (2/3 min.), 3/32-in. (1 min.), and 4/32-in.

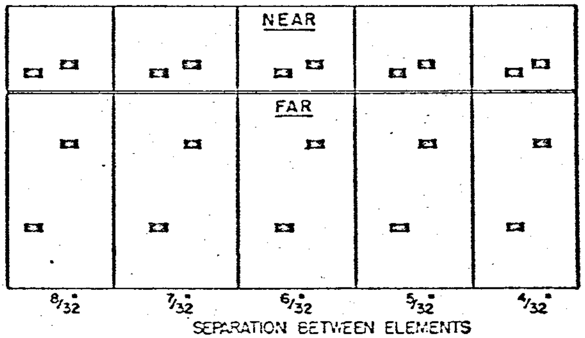

FIG. 7. SCALE-DRAWING OF THE BASIC ELEMENTS USED IN THE VISUAL PATTERNS

(1 1/3 min.) differences also were used. Sixteen presentations of each step-difference of a given pattern were made in random order in one session. Two sessions were given on each pattern, yielding 32 judgments by each *O* for each of the one-, two-, three-, and four-step differences.

At the beginning of each session, the series of five variations of the pattern were shown simultaneously to the *O* that he might see the differences to be discriminated. The cards then were removed and thereafter presented singly. Each card was exposed for 5 sec. with about 0.5 sec. between exposures in the same *ABX* series. The *O* was instructed to guess if he could not tell the difference between the patterns.

Four *O*s were used, all of whom had normal visual acuity of about 20/15 as measured by the Snellen Letter Chart. The order in which the patterns were presented was randomized except that each *O* completed all the patterns before repeating them the second time.

Results. The average results of the four *S*s are presented in Fig. 8 with the percentage of correct responses plotted as a function of the visual angle to be discriminated. The effect of the vertical spatial separation on the discriminative functions is shown separately for each pattern, the near-

separation always resulting in better discrimination than the far. The same superiority of near over far was found in the curves of each individual, although there were individual differences in the over-all level of discrimination.

The curves for the basic elements presented alone show that a difference of 1/3 min. of visual angle could not be discriminated as different by the *O*s. With larger visual angles, performance improves, the 75%

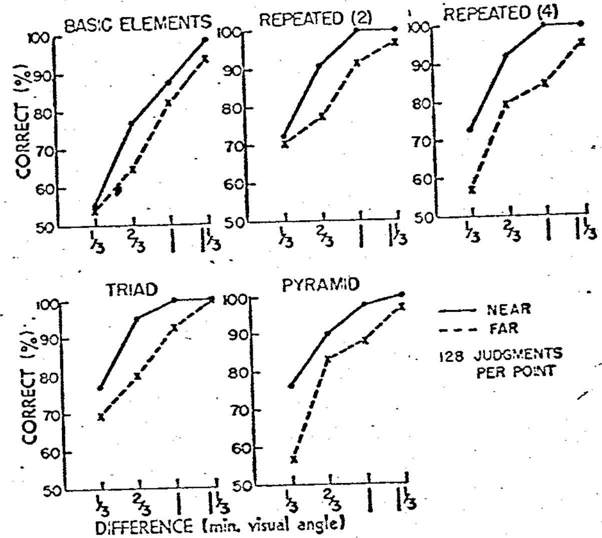

FIG. 8. RESULTS FOR THE TWO DEGREES OF VERTICAL SEPARATION IN EACH OF THE PATTERNS

point falling at about 2/3 min. for the near-condition and at about 5/6 min. for the far-condition.

Further inspection of Fig. 8 shows that, both in the near- and far-condition, repetition improves discrimination, although four repetitions are no more helpful than two. The triad and the pyramid yield better discrimination than the basic elements, although the pyramid is no better than the triad.

DISCUSSION

The effect of the frequency-separation of tones upon perception of sound-patterns, found by Miller and Heise, has been reproduced here under quite different conditions. It will be remembered that Miller and

Heise found a large frequency-separation between one of the tones in a pattern caused it to 'pop' out of the pattern.[9] It was pointed out previously that this effect should manifest itself in poorer temporal discrimination because the time-relations between an 'isolated' tone and the rest of the pattern should be more difficult to judge than the relations between tones all of which are in a single pattern. This expectation has been confirmed: the results show that temporal discrimination is poorer for the tone separated from the rest of the pattern by a greater difference in frequency.

A comparison of the results of the visual experiment with those from the auditory one reveals that one of the two major effects studied, that of the separation between elements, yielded the same result in both modalities. For both vision and audition, increasing the separation between the elements resulted in poorer discrimination under every condition. This then is direct support for the supposition that the spectrographic transform preserves the organization of the stimuli.

When the effect of the second major variable is considered, the results are not so definitive. One point is in complete agreement; patterns formed by repetitions of the basic elements improve discrimination in both senses, the differences between the repeated patterns and the basic elements alone being highly significant. Furthermore, it may be concluded that, while there is an advantage to be gained by repeating the basic elements once, there is no further advantage in increasing the number of repetitions. This is true for both vision and audition.

Further consideration of the effects of pattern-types, however, reveals an interesting discrepancy. The auditory triad yielded the poorest discriminative functions, while the visual triad was superior to the other patterns. Again, the basic elements ranked second to the repeated patterns in audition, while in vision this pattern was consistently the poorest.

This result indicates that the importance of some factor has been underestimated in the visual system. A possible consideration is that symmetry of pattern plays an important part in visual discrimination but not in auditory. Evidence for this supposition is provided by the *O*s who participated in the visual work. Each of them reported individually and spontaneously that judgments of the repeated patterns, under the far-condition, were impossible by trying to look at all of the elements. They used instead only the first three elements and made their judgments on this basis, the problem now being one of symmetry, and identical to the far triad, except that they had to ignore the unwanted elements. Further study is required to determine whether a comparable factor can be found in audition by varying the conversion-factors on the axes of the spectrographic transform, or whether this is a factor specific to vision for which there is no transformation comparable in audition.

[9] Heise and Miller, *op. cit.*, this JOURNAL, 72.

SUMMARY

The effect of two variables on temporal discrimination in auditory patterns and on spatial discrimination in visual patterns was measured. The two variables were the type of pattern, or the manner in which the elements were combined, and the degree of separation between elements. The patterns were transformed between the two senses by making frequency in audition comparable to the vertical dimension of visual space, and time in audition comparable to the horizontal dimension of visual space.

Temporal discrimination in auditory patterns was better when the elements in the pattern were of relatively near frequencies than when they were widely separated. This outcome was in accord with the visual result of better spatial discrimination when the elements were close together in the vertical dimension than when they were far apart.

Temporal discrimination also varied with the type of pattern in which the discrimination was imbedded. One repetition of the basic elements improved discrimination, but there was no further increase with further repetitions. The same result was found for visual (spatial) discrimination. The additional elements provided in the non-repeated patterns, as the triad and pyramid, did not aid discrimination in audition, but did improve it in vision.